

On Bounds and Algorithms for Frequency Synchronization for Collaborative Communication Systems

Peter A. Parker, Patrick Mitran, Daniel W. Bliss, and Vahid Tarokh

Abstract— Cooperative diversity systems are wireless communication systems designed to exploit cooperation among users to mitigate the effects of multipath fading. In fairly general conditions, it has been shown that these systems can achieve the diversity order of an equivalent MISO channel and, if the node geometry permits, virtually the same outage probability can be achieved as that of the equivalent MISO channel for a wide range of applicable SNR. However, much of the prior analysis has been performed under the assumption of perfect timing and frequency offset synchronization. In this paper, we derive the estimation bounds and associated maximum likelihood estimators for frequency offset estimation in a cooperative communication system. We show the benefit of adaptively tuning the frequency of the relay node in order to reduce estimation error at the destination. We also derive an efficient estimation algorithm, based on the correlation sequence of the data, which has mean squared error close to the Cramér-Rao Bound.

I. INTRODUCTION

Collaborative communication systems employ cooperation among nodes in a wireless network to increase data throughput and robustness to signal fading. Much of the research done in this area has concentrated on information theoretic results, protocols, and coding while assuming perfect synchronization [1]–[6]. In this paper, we explore frequency synchronization of a collaborative system and provide estimation bounds and practical algorithms having performance close to the bounds.

In a collaborative system, nodes that would have remained silent during some period of time adapt to their surroundings and collaborate with the source and destination nodes. These systems, sometimes termed cooperative diversity systems, use distributed protocols to greatly improve performance over traditional point-to-point communication systems. One improvement to system performance comes in the form of added robustness to signal fading [1], [2]. An effective way to achieve robustness is to increase the spatial diversity by using multiple antennas as in a MIMO system [7], [8]. However, when considering a network of low-cost wireless devices, the size and cost of multiple antennas is prohibitive for these devices [9]. A way for low cost nodes to realize much of the benefit of a MIMO system is through collaborative (cooperative) diversity.

This work was sponsored by the Department of the Air Force under Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

Portions of this work are to appear in the Proceedings of the International Conference on Distributed Computing Systems, June, 2007.

In fact, in [1] it is shown that a collaborative system can have the same diversity order as an equivalent MISO system. Employing a collaborative protocol in a wireless network can also increase the overall throughput of the network. The use of relaying is a special case of network coding and as shown in [10], the capacity of a relay (or coded) network is greater than in a traditional point-to-point network.

To design a practical collaborative communication system, one of two methods may be used. The signal modulation and coding may be designed to be naturally robust to synchronization errors [11], or alternatively, the frequency and timing offsets are estimated and subsequently compensated [12]. We explore the second option in this paper. Algorithms and bounds for standard synchronization are found in [13]–[15]. The related case of a MIMO channel with multiple frequency offsets is treated in [16], [17]. In this paper, we provide more details and extend the results of [18]. We derive the transmission frequency the relay must use to optimally reduce the variance of the frequency estimator at the destination by minimizing the Cramér-Rao Bound (CRB) of the frequency estimators at each receive node. By using the CRB, our frequency selection algorithm is independent of algorithm choice. We also provide an efficient frequency estimation algorithm for the collaborative system.

In [12], Shin et. al. describes a specific protocol, which we use in this paper, for collaborative communication with synchronization among three nodes: a source, a relay, and a destination. The protocol is based on a two-phase transmission within each frame [1], [4], a listening phase and a cooperation phase. Within each phase there is a preamble containing synchronization signals. In the listening phase, the relay receives and decodes the source's message. During the cooperation phase, the relay re-encodes and transmits the message cooperatively with the source. This process is illustrated in Figure 1.

The synchronization algorithms in [12] are ad-hoc and meant only to serve as a proof-of-concept that synchronization is possible with collaborative systems. In this paper, we derive the CRB for optimal frequency offset estimation for the class of systems discussed above. We show there exists an optimal (with respect to minimizing the CRB) frequency of transmission for the relay node based on: 1) the accuracy of estimation during the listening phase and 2) the SNR of all node pairs. We derive the maximum-likelihood (ML) frequency estimators for each receive node. These estimators are asymptotically efficient, meaning they achieve the CRB at high signal-to-noise

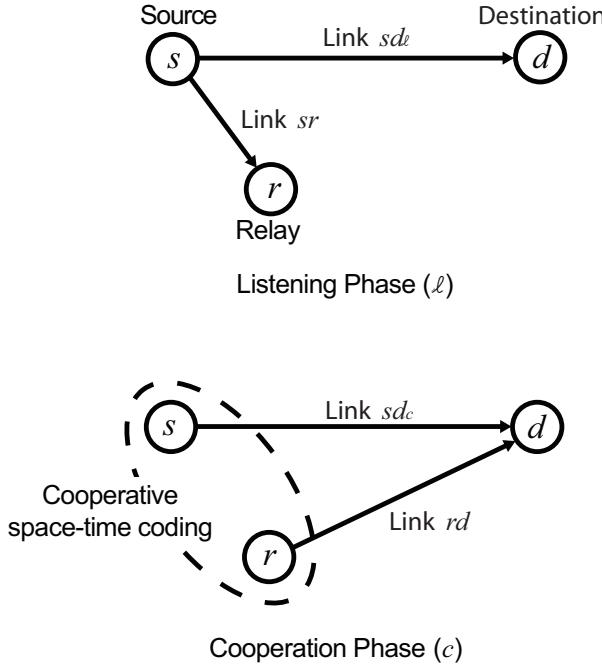


Fig. 1. Illustration of the two phases in a three node cooperative communication system.

ratio (SNR). However, the ML solution is computationally expensive and we therefore derive a practical correlation based estimation algorithm with performance close to the CRB. For the purposes of this paper, we assume a frequency selective fading model and that timing synchronization has been performed. Future papers will extend this work to include timing estimation and synchronization. We also assume all training sequences are constant modulus signals.

This paper is organized as follows, Section II outlines the mathematical model describing the signals involved in the frequency estimation portion of each phase. The CRB and ML estimators are derived in Sections III and IV for the listening and cooperation phases respectively. Section V provides some simulation results to illustrate the behavior and performance of frequency estimation in the three node relay system while Section VI shows the mean squared error (MSE) performance of each algorithm as compared with the CRB.

The following notation is used throughout: italic letters (x) represents scalar quantities, bold lowercase letters (\mathbf{x}) represent vectors, bold uppercase letters (\mathbf{A}) represent matrices, $(\cdot)^T$ denotes transpose, $(\cdot)^*$ denotes complex conjugation, $(\cdot)^H \triangleq (\cdot)^T$ denotes complex conjugate transpose, $\|\cdot\|$ denotes the 2-norm of a vector, $\Re(\cdot)$ denotes the real part of a complex number, $\mathbb{E}_{\mathbf{w}}(\cdot)$ denotes the expectation operator with respect to the random variable \mathbf{w} , $\mathcal{N}(\mu, \sigma^2)$ represents the Gaussian distribution with mean μ and variance σ^2 and $\mathcal{CN}(\mu, \sigma^2)$ represents the circularly symmetric complex Gaussian distribution, *i.e.*, where the real and imaginary parts are independent and identically distributed Gaussian random variables with variance $\sigma^2/2$.

II. SYSTEM MODEL

The system model is defined in this section. During each phase, a preamble consisting of a certain number of samples (N_ℓ for listening and N_c for cooperation) used for frequency synchronization. We assume the transmission channel is frequency selective with channel impulse response P samples long. Due to differences in local oscillator characteristics, the operating frequency of each node is slightly different. Let f_s denote the operating frequency of the source node and similar definitions for f_r and f_d . The notation sd is used to denote the source to destination link and likewise for sr and rd . As link sd is used in each phase, let sd_ℓ denote the link during the listening phase and sd_c be for the cooperation phase.

Each transmitted signal is received and converted to baseband for subsequent processing. During the listening phase, the baseband signal of link $a \in \{sd_\ell, sr\}$ is expressed as [15]

$$y_a[n] = e^{j2\pi f_a n} s_a[n] + w_b[n], \quad (1)$$

where n is the sample index, f_a is the frequency offset between the two nodes of link a normalized by the sample rate, $w_b[n]$ is the noise generated in the electronics of receiver $b \in \{d, r\}$ (destination or relay node respectively), and $s_a[n]$ is the combination of the known training signals ($\mathbf{x}_\ell = [x_\ell[0], \dots, x_\ell[N_\ell - 1]]^T$) and the effects of the frequency selective channel, given by

$$s_a[n] = \sum_{k=0}^{P-1} h_a[k] x_\ell[n-k].$$

In this equation, $h_a[n]$ are the samples of the channel response for link a and P is the duration of the channel response. We assume, for each link a , the length of the channel P is the same. Writing (1) in matrix form gives

$$\mathbf{y}_a = \mathbf{V}_{f_a} \mathbf{X}_\ell \mathbf{h}_a + \mathbf{w}_b \quad (2)$$

where $[\mathbf{V}_{f_a}]_{n,n} = e^{j2\pi f_a n}$ is a diagonal matrix and $[\mathbf{X}_\ell]_{i,k} = x_\ell[i-k]$ is a Toeplitz matrix where $x_\ell[k] = 0$ for $k < 0$ and $k \geq N_\ell$.

In the cooperation phase, the signal is defined as follows,

$$\mathbf{y}_c = \mathbf{V}_{f_{sd}} \mathbf{X}_{sd_c} \mathbf{h}_{sd_c} + \mathbf{V}_{f_{rd}} \mathbf{X}_{rd} \mathbf{h}_{rd} + \mathbf{w}_d, \quad (3)$$

where we assume the frequency f_{rd} is constant over both phases. For each receiver, b , the noise is assumed to be a zero-mean circularly symmetric complex Gaussian random vector

$$\mathbf{w}_b \sim \mathcal{CN}(\mathbf{0}, \sigma_b^2 \mathbf{I}).$$

In the general case, the frequency offsets between nodes can take on any values within the Doppler spread of the system plus the frequency differences of the local oscillators. We assume the maximum frequency offset is bounded and use this information to calculate the CRB and ML frequency estimators. In the remainder of the paper, we assume the nodes are stationary and thus the signals have no Doppler spread. A statistical model for the frequency offset is used as prior information to aid in frequency estimation. Let the operating frequency of each node $m \in \{r, s, d\}$ be modeled as

$$f_m = f_o + q_m,$$

where f_o is the mean operating frequency and q_m is a random variable with mean zero and variance σ_m^2 . We assume the random variables q_m are independent. For this paper, we also assume $\sigma_m^2 = \sigma_f^2$ for all nodes m , which is an appropriate model when considering a group of identical nodes cooperating together. The frequency offsets to be estimated are the difference between two of these independent random variables and thus the frequencies, f_a for $a \in \{sd, sr, rd\}$, have mean zero, variance $2\sigma_f^2$, and are correlated.

III. LISTENING PHASE

In the listening phase, the destination and the relay receive the same signal through two different channels. We drop the subscript a when considering only the single node-to-node link. To derive a good estimator for the frequency, it is useful to know the distribution of q_m . However, this is not known, so it is reasonable to design an estimator based on the “worst case” distribution constrained to the known statistics, *i.e.*, a mini-max estimator. As frequency estimation is inherently non-linear, an asymptotic analysis is performed in the high SNR regime (*i.e.*, $\text{SNR} \gg 1$). Under this assumption, the variance of a ML or maximum a posteriori (MAP) estimator is equal to the CRB. In the remainder of this section, we show that a Gaussian distribution with mean zero and variance σ_f^2 for q_m maximizes the CRB of the frequency estimate over all distributions with the same mean and variance. We then derive the MAP estimator of f .

A. Cramér-Rao Bound

The unknown parameters in the single node-pair model (1) are f (which is modeled as a random variable with mean zero and variance $2\sigma_f^2$) and \mathbf{h} .¹ The CRB is defined to be the diagonal entries of the inverse Fisher Information Matrix (FIM). When one or more parameters are random variables, the FIM is expressed in the following form [19]

$$\mathbf{J}_\theta = \mathbb{E}_f(\mathbf{J}_{\theta|f}) + \mathbf{J}_f, \quad (4)$$

where the expectation is taken over the random variable f ,

$$\mathbf{J}_{\theta|f} = -\mathbb{E}_w \left(\frac{\partial^2}{\partial \theta \partial \theta^T} L(\mathbf{y}|f) \right)$$

is the standard (non-random parameter) FIM, with expectation over the noise distribution, and $L(\mathbf{y}|, f) \propto \frac{-1}{\sigma^2} \|\mathbf{y} - \mathbf{V}_f \mathbf{X} \mathbf{h}\|^2$ is the log-likelihood of the data vector when the values of \mathbf{h} and f are held constant. The matrix \mathbf{J}_f is defined as follows:

$$\mathbf{J}_f = -\mathbb{E}_f \left(\frac{\partial^2}{\partial \theta \partial \theta^T} L(f) \right),$$

where $L(f) = \log p(f)$ and $p(f)$ is the distribution function of the random variable f . For the parameter vector $\theta^T = [f \ \mathbf{h}^T \ \bar{\mathbf{h}}^T]$, the FIM has the following form [20]

$$\mathbf{J}_{\theta|f} = \begin{bmatrix} \Delta & \Lambda & \bar{\Lambda} \\ \Lambda^T & \mathbf{0} & \Xi^T \\ \bar{\Lambda}^T & \Xi & \mathbf{0} \end{bmatrix}, \quad (5)$$

¹The parameter σ_f is considered known as it is a property of the receiver hardware. Also, the noise variance σ^2 is uncoupled with the other parameters and is estimated separately with no penalty.

where Δ is a scalar in this case. Let $[\mathbf{D}_\ell]_{nn} = 2n - 1 - N_\ell$ be a diagonal matrix such that

$$\frac{\partial}{\partial f} \mathbf{V}_f = j\pi \mathbf{D}_\ell \mathbf{V}_f. \quad (6)$$

The submatrices of (5) are computed as

$$\begin{aligned} \Delta &= \frac{2\pi^2}{\sigma^2} \|\mathbf{D}_\ell \mathbf{X} \mathbf{h}\|^2 \\ \Lambda &= \frac{-j\pi}{\sigma^2} \mathbf{h}^* \mathbf{X}^* \mathbf{D}_\ell \mathbf{X} \\ \Xi &= \frac{1}{\sigma^2} \mathbf{X}^* \mathbf{X}. \end{aligned}$$

None of these components depend on the random variable f and therefore the expectation in (4) goes away. The matrix \mathbf{J}_f is only non-zero in the first element and is

$$[\mathbf{J}_f]_{11} = -\mathbb{E} \left(\frac{\partial^2 L(f)}{\partial f^2} \right) \triangleq F_f, \quad (7)$$

where F_f is the Fisher information of the random variable f and $L(f)$ is the log-likelihood of f . The CRB for an estimator of f is then $[\mathbf{J}_\theta^{-1}]_{11}$, which can be calculated using the Shur complement² [21] to be

$$C_f = \left(\frac{2\pi^2}{\sigma^2} \|\mathbb{P}_X^\perp \mathbf{D}_\ell \mathbf{X} \mathbf{h}\|^2 + F_f \right)^{-1},$$

where $\mathbb{P}_X^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^*$ is the projection matrix onto the space orthogonal to the range of \mathbf{X} . As the Fisher information is a positive number, it is clear that, to find the worst case (maximum) CRB, F_f must be minimized. We use the following Lemma to show how this variable is minimized.

Lemma 1: Let $p_\sigma(\cdot)$ represent the family of distributions with mean zero and variance σ^2 . Let z be a random variable distributed as $p_\sigma(z)$. The minimum of the Fisher information of z , as defined in (7), over the family of distributions with variance σ^2 is achieved when

$$p_\sigma(z) = \mathcal{N}(0, \sigma^2).$$

Proof: Consider the following experiment: without any data, design an estimator \hat{z} for the random variable z . The log-likelihood in this case is $L(z) = \log p_\sigma(z)$. If $\hat{z} = 0$, then this estimator is unbiased and its variance is σ^2 . By the Cramér-Rao Theorem,

$$\text{var}(\hat{z}) \triangleq \sigma^2 \geq \frac{1}{F_z}.$$

Therefore, $F_z \geq \frac{1}{\sigma^2}$ with equality being achieved when $z \sim \mathcal{N}(0, \sigma^2)$. \square

By Lemma 1, the maximum CRB (over all distributions of f with variance $2\sigma_f^2$) is

$$C_f = \left(\frac{2\pi^2}{\sigma^2} \|\mathbb{P}_X^\perp \mathbf{D}_\ell \mathbf{X} \mathbf{h}\|^2 + \frac{1}{2\sigma_f^2} \right)^{-1}. \quad (8)$$

²The $\{1, 1\}$ block of a block matrix inverse is $[A^{-1}]_{11} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}$.

B. MAP Estimator of frequency

As a result of the preceding analysis, we use a Gaussian prior distribution on f to calculate the MAP estimator. This choice of prior represents the least informative prior of all distributions with variance $2\sigma_f^2$ and mean zero. For a particular channel gain \mathbf{h} , the log-likelihood of the data is

$$\begin{aligned} L(\mathbf{y}, f) &= \ln p(\mathbf{y}, f) = \ln p(\mathbf{y}|f) + \ln p(f) \\ &\propto \frac{-1}{\sigma^2} \|\mathbf{y} - \mathbf{V}_f \mathbf{X} \mathbf{h}\|^2 + \frac{1}{4\sigma_f^2} f^2. \end{aligned} \quad (9)$$

The apparent additional factor of two associated with σ_f^2 is due to the fact that f has a *real* Gaussian distribution as opposed to complex (as in the first term above). For any given frequency, the maximum of this expression over \mathbf{h} is achieved when

$$\hat{\mathbf{h}}(f) = (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{V}_f^* \mathbf{y}. \quad (10)$$

To find the MAP estimator of f , we substitute (10) into (9) and minimize the negative,

$$\hat{f} = \arg \min_f \left\{ \|\mathbf{P}_{\mathbf{X}}^\perp \mathbf{V}_f^* \mathbf{y}\|^2 + \frac{\sigma^2}{4\sigma_f^2} f^2 \right\}. \quad (11)$$

We note that as σ_f goes to infinity (no prior information), the estimator (11) is the standard ML frequency estimator [22].

IV. COOPERATION PHASE

In the cooperation phase, the destination node receives the superposition of signals coming from the source and relay. Each of these signals is transmitted with a slightly different frequency due to system imperfections. The purpose of this section is to derive a mini-max estimator for the two frequency offsets f_{sd} and f_{rd} . The estimator is mini-max in the sense that we design the (asymptotically) minimum variance estimator given that the prior distribution on the frequencies maximizes the estimator variance. We show there exists an optimal transmit frequency for the relay, which reduces the variance of frequency estimation at the destination.

As the relay has an estimate of f_{sr} (which is correlated with f_{sd} and f_{rd}) this information is useful in reducing the variance of the estimate at the destination. We assume the frequency transmitted from the relay is adjusted according to the following rule,

$$\begin{aligned} f_{r,Tx} &\triangleq f_r - \gamma \hat{f}_{sr} \\ &= f_r - \gamma (f_{sr} + e_{sr}) \end{aligned} \quad (12)$$

where γ is a parameter to be optimized and $e_{sr} \triangleq \hat{f}_{sr} - f_{sr}$ is the estimation error from the listening phase. We choose this rule as it is a linear function of the estimate and thus analytically tractable. When $\gamma = 0$, no frequency adjustment is made (e.g., when the estimate \hat{f}_{sr} provides no information about the source's frequency), and when $\gamma = 1$, the relay transmits its own estimate of the source's frequency (thus trusting the estimate to provide all of the information available about the source's frequency). We now express the frequency difference between the destination and the relay as

$$\begin{aligned} f_{rd} &= f_d - f_{r,Tx} \\ &= f_{sd} - (1 - \gamma) f_{sr} + \gamma e_{sr}. \end{aligned} \quad (13)$$

The two frequencies to be estimated at the destination node are f_{rd} and f_{sd} .

A. Covariance of frequencies

Before calculating the MAP estimator of f_{sd} and f_{rd} , we compute the least informative joint prior distribution. First, the covariance matrix of these random variables is found and then we show that the joint Gaussian distribution is the least informative prior.

To proceed, we calculate the covariance matrix of f_{sd} , f_{sr} , and e_{sr} . The mean of f_{sd} and f_{sr} are zero, $\mathbb{E}(f_{sd}^2) = \mathbb{E}(f_{sr}^2) = 2\sigma_f^2$ and $\mathbb{E}(f_{sd} f_{sr}) = \sigma_f^2$. Now consider $\mathbb{E}(e_{sr})$ (we show here that the MAP estimator derived above is asymptotically unbiased, i.e., $\mathbb{E}(e_{sr}) = 0$ for high SNR). Using the definition of e_{sr} and (11),

$$\begin{aligned} e_{sr} &= -f_{sr} + \xi \\ \xi &= \arg \min_f \left\{ \|\mathbf{V}_f \mathbf{P}_{\mathbf{X}_\ell}^\perp \mathbf{V}_f^* \mathbf{y}\|^2 + \frac{\sigma_r^2}{4\sigma_f^2} f^2 \right\}. \end{aligned}$$

By expressing the expectation as

$$\mathbb{E}(e_{sr}) = \mathbb{E}_{f_{sr}} (\mathbb{E}_{e_{sr}|f_{sr}} (\xi - f_{sr} | f_{sr})),$$

the conditional expectation $\mathbb{E}_{e_{sr}|f_{sr}} (\xi | f_{sr})$ needs to be calculated. Continuing the asymptotic analysis, for high SNR, we replace \mathbf{y} with its mean and obtain

$$\begin{aligned} \mathbb{E}_{e_{sr}|f_{sr}} (\xi | f_{sr}) &\approx \arg \min_f \left\{ \|\mathbf{V}_f \mathbf{P}_{\mathbf{X}_\ell}^\perp \mathbf{V}_f^* \mathbf{V}_{f_{sr}} \mathbf{X}_\ell \mathbf{h}_{sr}\|^2 \right. \\ &\quad \left. + \frac{\sigma_r^2}{4\sigma_f^2} f^2 \right\}, \end{aligned} \quad (14)$$

where the approximation is exact in the limit $\sigma_r^2 \rightarrow 0$. We perform the change of variables $\tilde{f}_{sr} = 0$ and $\tilde{f} = f - f_{sr}$, therefore, $\mathbf{V}_{\tilde{f}_{sr}} = \mathbf{I}$. The first term in (14) is

$$\mathbf{h}_{sr}^* \mathbf{X}_\ell^* \mathbf{V}_{\tilde{f}} \mathbf{P}_{\mathbf{X}_\ell}^\perp \mathbf{V}_{\tilde{f}}^* \mathbf{X}_\ell \mathbf{h}_{sr},$$

which is greater than or equal to zero and only equal to zero when $\tilde{f} = 0$ (i.e., $f = f_{sr}$). This function is thus locally convex about the point $f = f_{sr}$ and therefore locally quadratic. The second order Taylor series approximation is

$$\underbrace{\pi^2 \|\mathbf{P}_{\mathbf{X}_\ell}^\perp \mathbf{D}_\ell \mathbf{X}_\ell \mathbf{h}_{sr}\|^2}_{Q} \tilde{f}^2.$$

The value Q can be considered the effective signal power including all system and estimation gains. Returning to (14),

$$\begin{aligned} \mathbb{E}(\xi | f_{sr}) &\approx \arg \min_f \left\{ Q \cdot (f - f_{sr})^2 + \underbrace{\frac{\sigma_r^2}{4\sigma_f^2} f^2}_{K} \right\} \\ &= \frac{Q}{Q + K} f_{sr}. \end{aligned} \quad (15)$$

Completing the mean of e_{sr} ,

$$\mathbb{E}(e_{sr}) = \mathbb{E} \left(\frac{Q}{Q + K} f_{sr} - f_{sr} \right) = 0$$

because the mean of f_{sr} is zero and thus the estimator is asymptotically unbiased.

Continuing on with the covariance,

$$\mathbb{E}(f_{sr}e_{sr}) = \mathbb{E}(f_{sr}\mathbb{E}(e_{sr}|f_{sr})) = \frac{-2K}{Q+K}\sigma_f^2$$

and similarly $\mathbb{E}(f_{sd}e_{sr}) = \frac{-K}{Q+K}\sigma_f^2$ where K is defined in (15). Following a similar argument as above for $\mathbb{E}(e_{sr}^2)$ yields the result that the variance of e_{sr} is $\frac{2K}{Q+K}\sigma_f^2$, which is equal to the CRB in (8). Thus (11) is an asymptotically efficient estimate of the frequency. In summary,

$$\text{Cov}(f_{sd}, f_{sr}, e_{sr}) = \sigma_f^2 \begin{bmatrix} 2 & 1 & \frac{-K}{Q+K} \\ 1 & 2 & \frac{-2K}{Q+K} \\ \frac{-K}{Q+K} & \frac{-2K}{Q+K} & \frac{2K}{Q+K} \end{bmatrix}.$$

With this covariance matrix calculated, the covariance of f_{sd} and f_{rd} is

$$\mathbf{R}_{f_{sd}, f_{rd}} \triangleq \sigma_f^2 \begin{bmatrix} 2 & \frac{(1+\gamma)Q+K}{Q+K} \\ \frac{(1+\gamma)Q+K}{Q+K} & 2\frac{(1-\gamma+\gamma^2)Q+K}{Q+K} \end{bmatrix}. \quad (16)$$

B. Cramér-Rao Bound in Cooperative Phase

Recall the signal models for the cooperation phase (3) and the listening phase (2) as well as the relation between the two frequencies to be estimated f_{rd} and f_{sd} (13). The unknown parameters are f_{sd} , f_{rd} , \mathbf{h}_{sd_c} , \mathbf{h}_{rd} , and \mathbf{h}_{sd_ℓ} . For compactness, define $\mathbf{f} = [f_{sd} \ f_{rd}]^T$. The deterministic FIM ($\mathbf{J}_{\theta|\mathbf{f}}$) is a $(2+6P) \times (2+6P)$ matrix with the structure of (5) where Δ is 2×2 . Given the frequency random variables, the distributions of \mathbf{y}_c and \mathbf{y}_{sd_ℓ} are independent and the joint distribution is written as

$$p(\mathbf{y}_c, \mathbf{y}_{sd_\ell}, \mathbf{f}) = p(\mathbf{y}_c|\mathbf{f})p(\mathbf{y}_{sd_\ell}|\mathbf{f})p(\mathbf{f})$$

and the FIM is written as

$$\mathbf{J}_\theta = \mathbf{J}_{\theta|\mathbf{f}}(\mathbf{y}_c) + \mathbf{J}_{\theta|\mathbf{f}}(\mathbf{y}_{sd_\ell}) + \mathbf{J}_\mathbf{f}.$$

The blocks of the matrix $\mathbf{J}_{\theta|\mathbf{f}}(\mathbf{y}_c)$ are

$$\begin{aligned} \Delta_{11,c} &= \frac{2\pi^2}{\sigma_d^2} \|\mathbf{D}_c \mathbf{X}_{sd_c} \mathbf{h}_{sd_c}\|^2 \\ \Delta_{22,c} &= \frac{2\pi^2}{\sigma_d^2} \|\mathbf{D}_c \mathbf{X}_{rd} \mathbf{h}_{rd}\|^2 \\ \Delta_{12,c} &= \Delta_{21,c} = \frac{2\pi^2}{\sigma_d^2} \Re \{ \mathbf{h}_{sd_c}^* \mathbf{X}_{sd_c}^* \mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}} \mathbf{D}_c^2 \mathbf{X}_{rd} \mathbf{h}_{rd} \} \\ \Xi_{11,c} &= \frac{1}{\sigma_d^2} \mathbf{X}_{sd_c}^* \mathbf{X}_{sd_c} \\ \Xi_{22,c} &= \frac{1}{\sigma_d^2} \mathbf{X}_{rd}^* \mathbf{X}_{rd} \\ \Xi_{12,c} &= \Xi_{21,c}^* = \frac{1}{\sigma_d^2} \mathbf{X}_{sd_c}^* \mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}} \mathbf{X}_{rd} \\ \Lambda_{11,c} &= \frac{-j\pi}{\sigma_d^2} \mathbf{h}_{sd_c}^* \mathbf{X}_{sd_c}^* \mathbf{D}_c \mathbf{X}_{sd_c} \\ \Lambda_{22,c} &= \frac{-j\pi}{\sigma_d^2} \mathbf{h}_{rd}^* \mathbf{X}_{rd}^* \mathbf{D}_c \mathbf{X}_{rd} \\ \Lambda_{12,c} &= \frac{-j\pi}{\sigma_d^2} \mathbf{h}_{sd_c}^* \mathbf{X}_{sd_c}^* \mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}} \mathbf{D}_c \mathbf{X}_{rd} \end{aligned}$$

$$\Lambda_{21,c} = \frac{-j\pi}{\sigma_d^2} \mathbf{h}_{rd}^* \mathbf{X}_{rd}^* \mathbf{V}_{f_{rd}}^* \mathbf{V}_{f_{sd}} \mathbf{D}_c \mathbf{X}_{sd_c},$$

and zero for terms not listed. The diagonal matrix \mathbf{D}_c is defined similar to \mathbf{D}_ℓ in (6) with N_c replacing N_ℓ .

For data obtained during the listening phase, the matrix $\mathbf{J}_{\theta|\mathbf{f}}(\mathbf{y}_{sd_\ell})$ is

$$\begin{aligned} \Delta_{11,\ell} &= \frac{2\pi^2}{\sigma^2} \|\mathbf{D}_\ell \mathbf{X}_\ell \mathbf{h}_{sd_\ell}\|^2 \\ \Xi_{33,\ell} &= \frac{1}{\sigma^2} \mathbf{X}_\ell^* \mathbf{X}_\ell \\ \Lambda_{13,\ell} &= \frac{-j\pi^2}{\sigma} \mathbf{h}_{sd_\ell}^* \mathbf{X}_\ell^* \mathbf{D}_\ell \mathbf{X}_\ell \end{aligned}$$

and zero for terms not listed.

To calculate $\mathbb{E}(\mathbf{J}_{\theta|\mathbf{f}})$, note that only the (1,2) and (2,1) cross terms of the submatrices above (*i.e.*, Δ_{12} , $\Xi_{1,2}$, Λ_{12}, \dots) are dependent on the frequencies. In each case, the dependency is of the form $\mathbf{A} \mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}} \mathbf{B}$ where \mathbf{A} and \mathbf{B} are deterministic matrices or vectors. Looking at the n^{th} term of $\mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}}$,

$$\mathbb{E}([\mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}}]_{nn}) = \mathbb{E}(e^{j\pi d_n(f_{rd} - f_{sd})})$$

where $d_n = 2n - 1 - N_c$. This expectation is just the characteristic function of the random variable $f_{rd} - f_{sd}$ evaluated at πd_n (denoted as $\Phi_{f_{rd}-f_{sd}}(\pi d_n)$). Let $[\mathbf{M}]_{nn} = \Phi_{f_{rd}-f_{sd}}(\pi d_n)$ be a diagonal matrix, then we replace $\mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}}$ with \mathbf{M} in all cross terms of the FIM blocks. The FIM is then expressed as

$$\text{FIM} = \mathbb{E}(\mathbf{J}_{\theta|\mathbf{f}}) + \mathbf{J}_\mathbf{f} \quad (17)$$

where $\mathbf{J}_\mathbf{f}$ is nonzero only in the upper left 2×2 block and this block is equal to $\mathbf{F}_\mathbf{f}$, the Fisher information matrix of f_{sd} and f_{rd} . Using the Shur complement of the upper left 2×2 block of (17), the CRB for the frequencies are the diagonal entries of

$$\mathbf{C}_\mathbf{f} = (\Delta - 2\Re\{\Lambda\Xi^{-1}\Lambda^*\} + \mathbf{F}_\mathbf{f})^{-1}. \quad (18)$$

In the sequel, we desire to make conclusions about the performance of the collaborative system based on the derived bounds. As the absolute phase of the signal at each node is hard to control and cannot be relied on to remain stable over time, we find the worst case CRB and use this in the subsequent discussion. That is, for $\mathbf{h}_a = \tilde{\mathbf{h}}_a e^{j\phi}$, find ϕ maximizing the CRB (18). The resulting expression is

$$\mathbf{C}_{\mathbf{f},\max} = (\tilde{\Delta} - 2\text{abs}\{\Lambda\Xi^{-1}\Lambda^*\} + \mathbf{F}_\mathbf{f})^{-1}, \quad (19)$$

where $\tilde{\Delta}_{ii} = \Delta_{ii}$ and

$$\tilde{\Delta}_{12} = \tilde{\Delta}_{21} = \frac{-2\pi^2}{\sigma_d^2} \text{abs} \{ \mathbf{h}_{sd_c}^* \mathbf{X}_{sd_c}^* \mathbf{M} \mathbf{D}_c^2 \mathbf{X}_{rd} \mathbf{h}_{rd} \}.$$

Effectively, the phase ϕ is chosen to maximize magnitude of the off-diagonals of the matrix to be inverted in (19), which in turn maximizes the diagonals of the inverse (the negative signs are chosen for the off-diagonal terms because the FIM of the prior distribution, as calculated in the next section, also has negative off-diagonal terms).

C. Distribution of Frequencies

We now desire to find the distribution of f_{sd} and f_{rd} , which maximizes the CRB for a given frequency covariance \mathbf{R}_f (16). In order to do this, we assume the training sequences are chosen to provide near optimal performance. Examining (19), an ideal set of training sequences would zero out the off-diagonal terms in $\hat{\Delta}$ and also zero out the $(\Lambda \mathbf{E}^{-1} \Lambda^*)$ term. Thus for any constant modulus training sequences, the best CRB is

$$\mathbf{C}_{f,\text{opt}} = (\Delta_{\text{opt}} - \mathbf{F}_f)^{-1} \quad (20)$$

where $\Delta_{\text{opt}} = \text{diag}\{\Delta\}$. We show in Section V, by simulation, sequences exist where (19) is close to (20). Under the assumption of a good set of sequences, the dependence on the distribution of f_{sd} and f_{rd} enters only through \mathbf{F}_f . We use the following lemma to find the distribution maximizing the CRB.

Lemma 2: For \mathbf{A} , \mathbf{B} , and \mathbf{C} positive definite Hermitian matrices, if $\mathbf{B} > \mathbf{C}$ (i.e., $\mathbf{B} - \mathbf{C}$ is positive definite), then $(\mathbf{A} + \mathbf{C})^{-1} - (\mathbf{A} + \mathbf{B})^{-1}$ has positive diagonal entries.

Proof: By assumption, $(\mathbf{A} + \mathbf{B}) > (\mathbf{A} + \mathbf{C})$, which implies $(\mathbf{A} + \mathbf{C})^{-1} > (\mathbf{A} + \mathbf{B})^{-1}$. Thus the difference of the matrices is positive definite Hermitian and therefore has positive diagonal elements. \square

To maximize the diagonal elements of the CRB (20), Lemma 2 implies \mathbf{F}_f is as small as possible. Using an argument similar to the scalar case of Lemma 1, the Gaussian distribution satisfies this requirement and $\mathbf{F}_f = \mathbf{R}_f^{-1}$. The assumptions that f_s , f_r , f_d and the estimation error from the listening phase e_{sr} are jointly Gaussian is therefore the least informative prior given the specified variances and correlations.

D. Optimal γ

With the aim of deriving a mini-max estimator, we desire to choose γ in (12) to minimize the trace of \mathbf{C}_f (20). As this expression is not intuitive, it is helpful to consider a flat fading model. For flat fading, $P = 1$ and the terms in the optimal CRB (20) are

$$\Delta_{\text{opt}} = \begin{bmatrix} \eta_c S_{sd_c} + \eta_\ell S_{sd_\ell} & \mathbf{0} \\ \mathbf{0} & \eta_c S_{rd} \end{bmatrix}$$

and

$$\mathbf{R}_f = \begin{bmatrix} 2 & \frac{2\eta_\ell(1+\gamma)S_{sr}+1/\sigma_f^2}{2\eta_\ell S_{sr}+1/\sigma_f^2} \\ \frac{2\eta_\ell(1+\gamma)S_{sr}+1/\sigma_f^2}{2\eta_\ell S_{sr}+1/\sigma_f^2} & 2\frac{2\eta_\ell(1-\gamma+\gamma^2)S_{sr}+1/\sigma_f^2}{2\eta_\ell S_{sr}+1/\sigma_f^2} \end{bmatrix} \sigma_f^2.$$

where $\eta_\ell = \frac{2}{3}\pi^2 N_\ell (N_\ell^2 - 1)$ (similarly for η_c) and $S_{rd} = \frac{|h_{rd}|^2}{\sigma_d^2}$ is the signal to noise ratio of the source-relay link (similarly for S_{sd_c} , S_{sd_ℓ} , and S_{sr}).

An exact calculation of the optimal γ leads to a long, complicated expression that depends on the SNR of each link and the variance of the frequency oscillators. The expression is omitted here as it gives no insight into the problem. Later, we show there is minimal loss when γ is always set to 1. To gain some insight into the behavior of γ , consider two limiting cases for γ_{opt} : $\sigma_f \rightarrow \infty$ and $\sigma_f \rightarrow 0$.

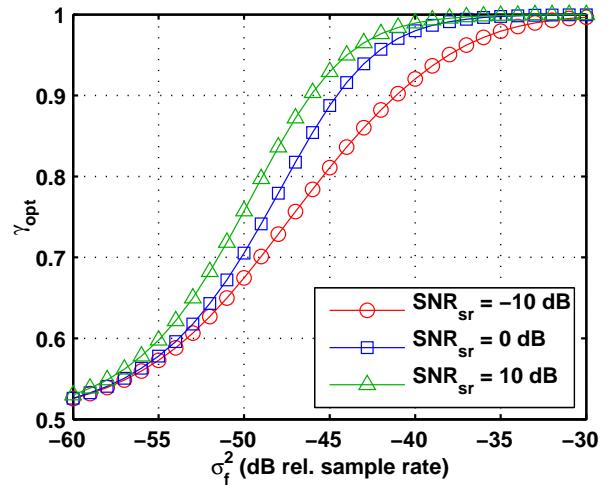


Fig. 2. Plot of optimal γ as a function of modeled frequency variation. Three curves are shown for different values of gain between the source and relay. The SNR of the source-destination and relay-destination node pairs are held constant at 0 dB.

1) *Large σ_f , or no prior information:* By taking the limit of the expression for γ_{opt} as $\sigma_f \rightarrow \infty$, it can be shown that $\gamma_{\text{opt}} \rightarrow 1$. In this case, the relay transmits at a frequency equal to its estimate of the source frequency. By choosing this transmit frequency, the operating frequency of the relay f_r is removed from the estimation procedure as it contains no information about the source-destination frequency.

2) *Small σ_f :* When $\sigma_f \rightarrow 0$ (or when $1/\sigma_f^2$ is much larger than any of the link SNRs perhaps due to poor channel SNRs), the CRB is minimized when $\gamma = 1/2$. By looking at the MAP frequency estimator (11) for this limiting case, the frequency estimate is zero. Therefore, no matter what γ is chosen, the relay just transmits at its own frequency. When σ_f is small (but not zero), there is still some information in the frequency estimate about the source frequency (besides the information from the local oscillator model), and by choosing $\gamma \approx 1/2$, both sources of information are used to select the best transmit frequency.

As an example of the function γ_{opt} , Figure 2 shows plots of several curves of γ_{opt} versus σ_f^2 . The length of the training signal is $N_\ell = N_c = 16$ and the SNR of the source-destination link is -3 dB (combining the listening and cooperation phases, the effective SNR is 0 dB). The SNR from relay to destination, S_{rd} , is also 0 dB and there is one curve each for $S_{sr} \in \{-10 \text{ dB}, 0 \text{ dB}, 10 \text{ dB}\}$. For each curve, the transition from $\gamma_{\text{opt}} = 1$ to $\gamma_{\text{opt}} = 1/2$ appears to occur roughly when $\sigma_f^2 \approx (\eta_c S_{sd_c} + \eta_\ell S_{sd_\ell})$ or $\sigma_f^2 \approx \eta_c S_{rd}$. These values of σ_f^2 are significant because, for example, when $\sigma_f^2 < \eta_\ell S_{sd_\ell}$ (left half of the plot), the assumed prior knowledge of frequency has more weight than the data, whereas when $\sigma_f^2 > \eta_\ell S_{sd_\ell}$, the information in the data is more important than the prior model.

E. MAP Estimator of f_{sd} and f_{rd}

To calculate the MAP estimate of f_{sd} and f_{rd} at the destination node during the cooperation phase, the covariance

between these two random variables (16) is needed. Therefore, the values of Q and K need to be forwarded to the destination node. The log-likelihood of the data at the destination is

$$\begin{aligned} L(\mathbf{y}_c, \mathbf{y}_{sd\ell}, \mathbf{f}) &= \ln p(\mathbf{y}_c|\mathbf{f}) + \ln p(\mathbf{y}_{sd\ell}|\mathbf{f}) + p(\mathbf{f}) \\ &\propto \frac{-1}{\sigma_d^2} \|\mathbf{y}_{sd\ell} - \mathbf{V}_{f_{sd}} \mathbf{X}_\ell \mathbf{h}_{sd\ell}\|^2 \\ &\quad - \frac{1}{\sigma_d^2} \|\mathbf{y}_c - \mathbb{X}(\mathbf{f}) \mathbf{g}\|^2 - \frac{1}{2} \mathbf{f}^T \mathbf{R}_f^{-1} \mathbf{f}, \end{aligned} \quad (21)$$

where $\mathbb{X}(\mathbf{f}) \triangleq [\mathbf{V}_{f_{sd}} \mathbf{X}_{sd_c} \mathbf{V}_{f_{rd}} \mathbf{X}_{rd}]$ and $\mathbf{g}^T = [\mathbf{h}_{sd_c}^T \mathbf{h}_{rd}^T]$. As before, choose estimates for \mathbf{g} and $\mathbf{h}_{sd\ell}$ to maximize the likelihood for any given frequency pair,

$$\begin{aligned} \hat{\mathbf{g}}(\mathbf{f}) &= (\mathbb{X}^* \mathbb{X})^{-1} \mathbb{X}^* \mathbf{y}_c \\ \hat{\mathbf{h}}_{sd\ell}(f_{sd}) &= (\mathbf{X}_\ell^* \mathbf{X}_\ell)^{-1} \mathbf{X}_\ell^* \mathbf{V}_{f_{sd}}^* \mathbf{y}_{sd\ell}. \end{aligned}$$

Substituting these estimates into (21) and minimizing the negative to obtain the MAP frequency estimator

$$\begin{aligned} \hat{\mathbf{f}} &= \arg \min_{\mathbf{f}} \left\{ \|\mathbb{P}_{\mathbb{X}(\mathbf{f})}^{\perp} \mathbf{y}_c\|^2 + \|\mathbb{P}_{\mathbf{X}_\ell}^{\perp} \mathbf{V}_{f_{sd}} \mathbf{y}_{sd\ell}\|^2 + \right. \\ &\quad \left. + \frac{\sigma_d^2}{2} \mathbf{f}^T \mathbf{R}_f^{-1} \mathbf{f} \right\}. \end{aligned} \quad (22)$$

We note the special case of $\sigma_f \rightarrow \infty$ (which implies $\gamma_{opt} = 1$). For $\gamma = 1$, the covariance (16) needed in the MAP estimator simplifies to

$$\mathbf{R}_{f_{sd}, f_{rd}} = \sigma_f^2 \begin{bmatrix} 2 & \frac{2Q+K}{Q+K} \\ \frac{2Q+K}{Q+K} & 2 \end{bmatrix},$$

which has a finite inverse when $\sigma_f < \infty$. However, when $\sigma_f \rightarrow \infty$, we evaluate the limit of \mathbf{R}_f^{-1} resulting in

$$\lim_{\sigma_f \rightarrow \infty} \mathbf{R}_f = \zeta \zeta^T \frac{2Q}{\sigma_r^2} = \zeta \zeta^T \underbrace{\frac{2\pi^2}{\sigma_r^2} \|\mathbb{P}_{\mathbf{X}_\ell}^{\perp} \mathbf{D}_\ell \mathbf{X}_\ell \mathbf{h}_{sr}\|^2}_{C_{f_{sr}}}$$

where $\zeta^T = [1 \ -1]$ and $C_{f_{sr}}$ is the CRB of the frequency in the source-relay link (8) with $\sigma_f = \infty$. The penalty term (last term) of the MAP estimator (22) simplifies to

$$\frac{\sigma_d^2}{2} \mathbf{f}^T \mathbf{R}_f^{-1} \mathbf{f} \rightarrow \frac{\sigma_d^2}{2C_{f_{sr}}} (f_{sd} - f_{rd})^2.$$

Thus the penalty term is a quadratic of the frequency difference term normalized by the ratio of error variances (noise power over frequency estimation error variance).

V. SIMULATIONS

In the previous section, we showed the optimal γ for extreme values of σ_f is either 1 or $1/2$ and when γ_{opt} approaches $1/2$, its effect is small because the frequency adjustment is going toward zero. In this section, we show by simulation, the penalty for choosing $\gamma = 1$ instead of $\gamma = \gamma_{opt}$ is usually limited to a few tenths of a decibel. Thus, near optimal performance is achieved without communicating any of the link SNRs back to the relay for calculation of γ_{opt} . We also show the existence of training sequences where (19) is close to (20). Finally, we show the benefit of letting the

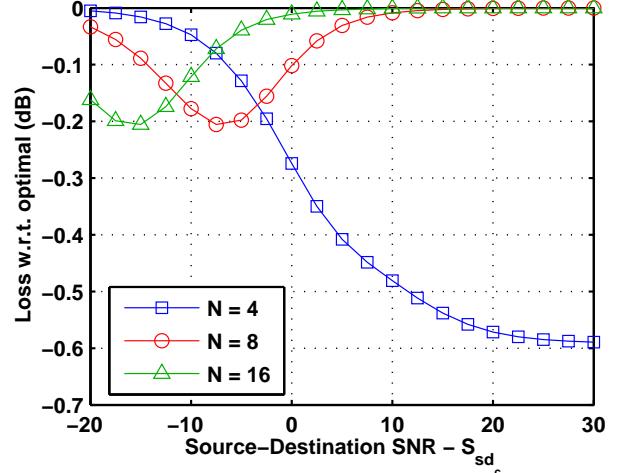


Fig. 3. Plot of the loss in performance caused by binary training sequence as opposed to an arbitrary sequence, and when choosing $\gamma = 1$ versus $\gamma = \gamma_{opt}$. Relay-destination and source-destination SNRs are the same and source-relay SNR is 10 dB higher.

relay set its transmit frequency based on information received during the listening phase.

We simulate a three node system in a frequency flat environment. In all simulations, we use the SNR of the sd link (assuming $S_{sd_c} = S_{sd\ell}$) as a reference value. The following configuration is considered: let $S_{rd} = S_{sd_c}$ and then vary the link SNR of the source-relay link relative to S_{sd_c} . Let $N_\ell = N_c$. The prior distribution for the operating frequency we assume is Gaussian with a variance of -40 dB relative to the sample rate (e.g., a 2 parts-per-million variance of a local oscillator at 900 MHz with 4.5 MHz sample rate [12]).

For flat fading channels and constant modulus training sequences, it is sufficient to choose $\mathbf{x}_\ell = \mathbf{1}$ (the vector of all ones) and $\mathbf{x}_{sd_c} = \mathbf{1}$. A search is performed to find \mathbf{x}_{rd} which minimizes the CRB (19). For values of $N_c \in \{4, 8, 16\}$ an exhaustive search over all binary sequences is performed (results hold independent of choice between $\gamma = \gamma_{opt}$ or $\gamma = 1$) and for values of $N_c > 16$, a randomized search over binary sequences is performed. For each value of N_c (up to 128) the optimal sequence for \mathbf{x}_{rd} has the following structure:

Sequence Design: Let $\mathbf{a}_1 = [1, -1]^T$ and

$$\mathbf{a}_n^T = [\mathbf{a}_{n-1}^T, -\mathbf{a}_{n-1}^T]$$

where \mathbf{a}_n is length 2^n and is the last column of a Sylvester matrix. Then the length $N_c = 2^n$ optimal sequence is

$$\mathbf{x}_{rd, opt} = \begin{bmatrix} \mathbf{a}_{n-1} \\ -\mathbf{J} \mathbf{a}_{n-1} \end{bmatrix}$$

where \mathbf{J} is the exchange matrix which reverses the order of elements in the vector it multiplies.

For the configuration described above, and with $S_{sr} = 10 S_{sd_c}$, Figure 3 shows the difference between the best possible CRB (20) (for any constant modulus sequence and $\gamma = \gamma_{opt}$) and the worst case CRB (19) using the binary sequence shown above and $\gamma = 1$. The 0.6 dB difference for $N_c = 4$ is primarily due to a non-optimal sequence

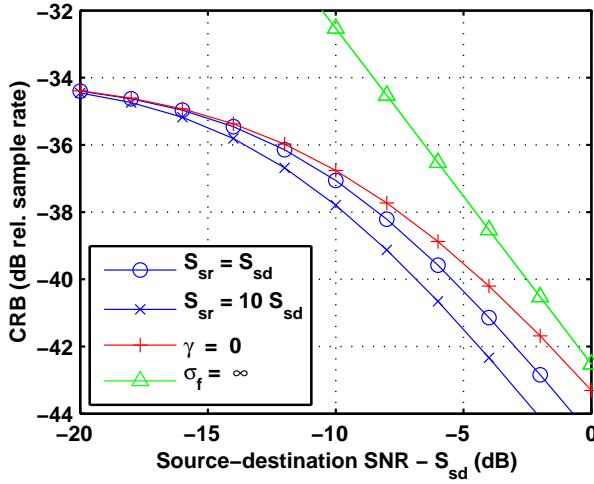


Fig. 4. Plot of the sum of Cramér-Rao Bounds for f_{sd} and f_{rd} . Circle and “x”-marks show bound when $\sigma_f^2 = -40$ dB and $\gamma = \gamma_{opt}$, plus marks show bound when $\gamma = 0$, and triangles show bound when $\gamma = 0$ and $\sigma_f = \infty$ (the standard frequency bound assuming no prior information). All curves are for a length 16 training sequence.

x_{rd} , whereas the 0.2 dB difference for other values of N_c is due to choosing $\gamma = 1$ instead of the optimal value. The loss in performance due to a non-optimal sequence decreases dramatically as N_c increases. These loss values are typical of other system configurations as well. The system behavior as a function of training sequence illustrates the fact that the CRB is insensitive to the selection of these sequences.

Figure 4 shows the sum of the CRB for the two frequencies estimated at the destination node as a function of S_{sd_c} . For this figure, the SNRs of the source-destination link (S_{sd}) and relay-destination link (S_{rd}) are the same. The circle and “x”-marks show the CRB when the SNR of the source-relay link S_{sr} is, respectively, the same as and 10 dB higher than S_{sd} . The plus marks show the CRB when $\gamma = 0$. The difference between the plus-marks and the circle and “x”-marks show the potential gain in estimation performance by changing the relay’s transmit frequency (greater benefit when the SNR is large). The triangles show the CRB when no prior information is used. This shows a great advantage of using a prior model when the SNR is low.

VI. SUB-OPTIMAL ALGORITHMS

The maximum-likelihood frequency estimator (22) requires a two-dimensional search over the frequency range of interest. As this is a computationally expensive approach to estimation, we compare the mean squared error (MSE) performance of more efficient, sub-optimal estimation algorithms and introduce a correlation based estimator as the best compromise between estimation performance and computational efficiency. In the remainder of this section, we describe the use of the one-dimensional ML algorithm as applied to the two signal case and the correlation algorithm for frequency estimation and compare their performance.

A. One-Dimensional ML

As a result of the choice of the training sequence, estimation of the two frequencies is nearly uncoupled. Therefore, performing two independent one-dimensional ML searches for the frequencies is approximately the same as performing the full two-dimensional ML search as required by the ML algorithm. Given the data vector \mathbf{y}_c , the one-dimensional ML estimates of the frequencies are

$$\hat{f}_{rd} = \arg \min_f \left\{ \|\mathbb{P}_{\mathbf{x}_{rd}}^\perp \mathbf{V}_f^* \mathbf{y}_c\|^2 + \frac{\sigma_d}{4\sigma_f} f^2 \right\} \quad (23)$$

$$\begin{aligned} \hat{f}_{sd} = \arg \min_f & \{ \|\mathbb{P}_{\mathbf{x}_{sd}}^\perp \mathbf{V}_f^* \mathbf{y}_{sd}\|^2 + \\ & + \|\mathbb{P}_{\mathbf{x}_{sd}}^\perp \mathbf{V}_f^* \mathbf{y}_c\|^2 + \frac{\sigma_d}{4\sigma_f} f^2 \}, \end{aligned} \quad (24)$$

which do not take the correlations between the frequencies into account. To improve the estimates (23) and (24), we assume the variance of each estimate meets the CRB assuming the prior information is uncorrelated for each frequency:

$$\tilde{\mathbf{C}}_f = (\Delta - 2\Re\{\Lambda\Xi^{-1}\Lambda^*\} + \text{diag}\{\mathbf{R}_f\}^{-1})^{-1},$$

where $\text{diag}\{\mathbf{R}_f\}$ is a diagonal matrix consisting of the diagonal entries of \mathbf{R}_f (zeroing out the other elements). This assumption is valid for high SNR and large N_c . Incorporating this knowledge with the prior information, the least squares estimates of the frequencies are

$$\begin{bmatrix} \hat{f}_{sd,ML1} \\ \hat{f}_{rd,ML1} \end{bmatrix} = \mathbf{R}_f (\mathbf{R}_f + \tilde{\mathbf{C}}_f)^{-1} \begin{bmatrix} \hat{f}_{sd} \\ \hat{f}_{rd} \end{bmatrix}. \quad (25)$$

B. Correlation Method

We first describe a standard correlation frequency estimation method as presented in [23] and then provide an extension to allow this algorithm to work in the presence of two signals with known training sequences. Assuming a single signal in the presence of flat fading

$$y[n] = e^{j2\pi f n} x[n] + w[n], \quad 1 \leq n \leq N.$$

The estimated autocorrelation sequence of $y[n]$ is

$$R[k] = \frac{1}{N-k} \sum_{i=k+1}^N (y[n]\bar{x}[n])(\bar{y}[i-k]x[i-k]).$$

The estimate of the frequency is calculated as

$$\hat{f} = \frac{1}{\pi(M+1)} \arg \left\{ \sum_{k=1}^M R[k] \right\} \quad (26)$$

where M is a design parameter and the frequency estimate is unambiguous if

$$|f| < \frac{1}{M+1}.$$

Therefore, M trades performance for estimation range. The performance of this algorithm (26) is shown in [23] to be close to the CRB when $M = N/2$. To ensure adequate estimation range, the maximum allowed value of M is 12 (corresponding to a range of five standard deviations away from the mean of the prior). To incorporate the known prior knowledge of the

frequency variance, the estimate (26) is adjusted according to the following rule

$$\hat{f}_p = \frac{2\sigma_f^2}{2\sigma_f^2 + c_f^2} \hat{f}$$

where c_f^2 is the CRB of the frequency estimate with no prior information. Let

$$\hat{f} = \rho(\mathbf{y}, \mathbf{x}, \sigma_f)$$

be a function that inputs the data vector \mathbf{y} , training vector \mathbf{x} , and prior information, and outputs the frequency estimate according to the above algorithm. This algorithm is used without modification during the listening phase to calculate the estimate $\hat{f}_{sr} = \rho(\mathbf{y}_{sr}, \mathbf{x}_{\ell}, \sigma_f)$.

For the cooperation phase, there are two signals present and the undesired signal acts as interference for the desired signal being estimated. The estimates provided by the correlation algorithm are

$$\tilde{f}_{sd,1} = \rho(\mathbf{y}_c, \mathbf{x}_{sd}, \sigma_f) \quad (27)$$

$$\tilde{f}_{rd,1} = \rho(\mathbf{y}_c, \mathbf{x}_{rd}, \sigma_f), \quad (28)$$

which exhibit a floor in MSE (see Figure 5). To improve the estimates, we project out the undesired signal in the following manner:

$$\tilde{\mathbf{y}}_{c,sd} = \mathbb{P}_{\mathbf{V}_{\tilde{f}_{rd,1}}}^\perp \mathbf{x}_{rd} \mathbf{y}_c$$

$$\tilde{\mathbf{y}}_{c,rd} = \mathbb{P}_{\mathbf{V}_{\tilde{f}_{sd,1}}}^\perp \mathbf{x}_{sd} \mathbf{y}_c,$$

where the frequency estimates in (27) and (28) are used to calculate the interference signal, which is projected out. The correlation algorithm is run a second time to find

$$\tilde{f}_{sd,2} = \rho(\tilde{\mathbf{y}}_{c,sd}, \mathbf{x}_{sd}, \sigma_f)$$

$$\tilde{f}_{rd,2} = \rho(\tilde{\mathbf{y}}_{c,rd}, \mathbf{x}_{rd}, \sigma_f).$$

The final frequency estimates, with all prior information accounted for, is calculated similarly to (25),

$$\begin{bmatrix} \hat{f}_{sd,corr} \\ \hat{f}_{rd,corr} \end{bmatrix} = \mathbf{R}_f (\mathbf{R}_f + \tilde{\mathbf{C}}_f)^{-1} \begin{bmatrix} \tilde{f}_{sd,2} \\ \tilde{f}_{rd,2} \end{bmatrix}.$$

Figure 5 shows the total MSE (summation of errors from \hat{f}_{sd} and \hat{f}_{rd}) of the correlation algorithm compared with the CRB for $N_c = 16$. The triangle markers denote the performance of the algorithm without any adaptation while the circle markers denote the performance of the adaptive two-step algorithm described above. For lower SNRs, the adaptive algorithm has about a 3 dB advantage while the performance difference is much greater at higher SNRs (above 15 dB). The performance of the adaptive algorithm is near optimal. The slight “bump” in performance of the two algorithms at $S_{sd} = -10$ dB SNR is caused by the interaction of the threshold region (the region where the MSE performance breaks away from the CRB) and the region dominated by prior information (where the algorithms converge to a -34 dB MSE relative to the sample rate).

For the same scenario, Figure 6 compares the three estimation algorithms: full (two-dimensional search) maximum likelihood (circles), one-dimensional ML (“x”-marks), and adaptive correlation (triangles). The mean squared error is compared with the CRB.

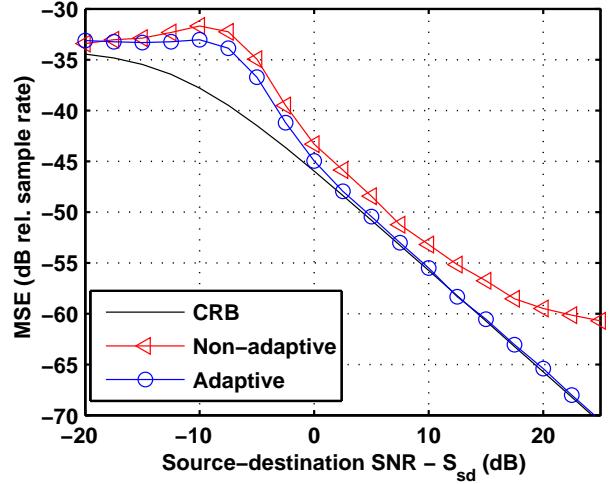


Fig. 5. Plot of mean squared error of non-adaptive (circles) and adaptive two-step (triangles) correlation algorithms. The mean squared error is compared with the CRB.

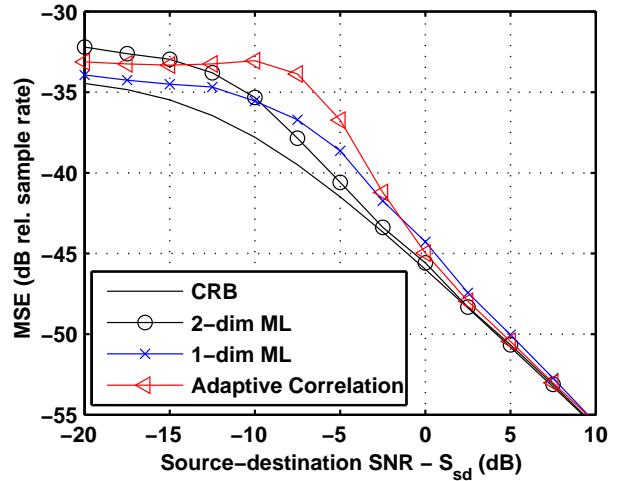


Fig. 6. Plot of mean squared error of full (two-dimensional search) ML (circles), one-dimensional ML (“x”-marks), and adaptive correlation (triangles). The mean squared error is compared with the CRB.

and the adaptive correlation algorithm (triangles). Each of these algorithms approaches the CRB asymptotically in SNR. The differences in behavior at lower SNRs is attributed to the different algorithms entering their threshold regions at different SNRs. A more detailed analysis of this region can be carried out using the methods of [24].

VII. CONCLUSIONS

In this paper, we have derived the Cramér-Rao bounds for frequency offset estimation in a three-node collaborative communication system. We have shown through simulation, the performance increase obtained by allowing the relay to change its transmitting frequency. We have also shown there exists an optimal transmit frequency for the relay node based on the other link SNRs and the assumed prior knowledge of the frequency offsets. However, there is only a small (tenths of decibels) penalty if the relay always transmits at its estimate of

the source frequency. Simulation results also demonstrate the existence of binary training sequences that result in very little loss as compared with an arbitrary constant modulus sequence. We also derived a computationally efficient correlation based estimation algorithm that has mean squared error performance close to the CRB.

REFERENCES

- [1] P. Mitran, H. Ochiai, and V. Tarokh, "Space-time diversity enhancements using collaborative communications," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2041–2057, June 2005.
- [2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity – part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.
- [3] ———, "User cooperation diversity – part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.
- [4] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [6] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: Space-time transmission and iterative decoding," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 362–371, Feb. 2004.
- [7] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, pp. 311–335, 1998.
- [8] I. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [9] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey of sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.
- [10] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: the relay case," in *INFOCOM 2002*, vol. 3, 2002, pp. 1577–1586.
- [11] X. Li, "Space-time coded multi-transmission among distributed transmitters without perfect synchronization," *IEEE Signal Process. Lett.*, vol. 11, no. 12, pp. 948–951, Dec. 2004.
- [12] O. Shin, A. Chan, H. T. Kung, and V. Tarokh, "Design of an OFDM cooperative space-time diversity system," *IEEE Trans. Veh. Technol.*, to appear.
- [13] J. van de Beek, M. Sandell, and P. O. Börjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1800–1805, July 1997.
- [14] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.
- [15] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580–1589, Sept. 2000.
- [16] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 602–613, Mar. 2003.
- [17] S. Ahmed, S. Lambotharan, A. Jakobsson, and J. A. Chambers, "MIMO frequency-selective channels with multiple frequency offsets: estimation and detection techniques," *IEE Proc. Commun.*, vol. 152, no. 4, pp. 489–494, Aug. 2005.
- [18] P. A. Parker, P. Mitran, D. W. Bliss, and V. Tarokh, "Adaptive frequency synchronization for collaborative communication systems," in *Int. Conf. on Distributed Computing Systems (ICDCS)*, Toronto, Canada, June 2007.
- [19] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Part 1*. Wiley Interscience, 1968.
- [20] S. T. Smith, "Statistical resolution limits and the complexified Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1597–1609, May 2005.
- [21] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.
- [22] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison Wesley, 1991.
- [23] M. Luise and R. Reggiannini, "Carrier frequency recovery in all-digital modems for burst-mode transmissions," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1169–1178, Feb./Mar./Apr. 1995.
- [24] C. D. Richmond, "Mean-squared error and threshold SNR prediction of maximum-likelihood signal parameter estimation with estimated colored noise covariances," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2146–2164, May 2006.